DOCUMENT RESUME

ED 295 124 CS 009 145

AUTHOR Dollerup, Cay; And Others
TITLE Vocabularies in the Reading Process.
PUB DATE 88
NOTE 34p.
PUB TYPE Reports - General (140)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Foreign Countries; Higher Education; *Language Tests; *Reading Processes; Reading Research; Reading Strategies; Secondary Education; *Vocabulary; Vocabulary Development; Word Recognition
IDENTIFIERS *Denmark

ABSTRACT

Using a study of 30 Danish freshman college students' vocabularies as a springboard, this paper explores and discusses a number of current assumptions about vocabularies in the mother tongue and in foreign language pedagogics. The paper contends that it is logical to assume that vocabularies in reading are fluid and depend on the text read, on the reading strategies employed, and on the words the readers feel they know. The paper argues that a reader's vocabulary is part of the process of reading: it is a function of the text and its contents, of the reader's reading strategies, and of the reader's more or less stable "word knowledge," with a constant interplay between these factors. This implies that a deliberate teaching of reading strategies in addition to other methods would be pedagogically sound. The major portion of the paper discusses in detail how current assumptions fall short of explaining the findings in the original study which led to the reading strategies theory. One illustration, two charts, two tables of data and two notes are included, and 22 references and two appendixes are attached. (MS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED295124

TITLE:

VOCABULARIES IN THE READING PROCESS

Authors:

Cay Dollerup, Ester Glahn, Carsten Rosenberg Hansen,

COPENHAGEN, DENMARK

ABSTRACT:

Using a study of Danish freshman undergraduates' vocabularies as a springboard, the paper explores and discusses a number of current assumptions about vocabularies in the mother tongue and in foreign language pedagogics. The conclusion is that as far as reading is concerned, a reader's vocabulary is part of the process of reading: it is a function of the texts and its contents, of the reader's reading strategies, and of the reader's more or less stable "word knowledge". In the reading of a specific text there is a constant interplay between these factors which implies that a vocabulary in reading is "fluid".

Pedagogically, this theory implies only that there should be a deliberate teaching of reading strategies in addition to other methods. Ontologically, the theory implies that passive vocabularies in reading exist instantaneously and are variable entities in each reading of a specific text.

In order to carry conviction, and better to discuss much thinking about vocabularies, the main bulk of the paper is a very detailed discussion of how current assumptions fall short of explaining the findings in the study which led us to the above theory. (CD)

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Cay Dollerup

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CS009145

VOCABULARIES IN THE READING PROCESS

CAY DOLLERUP
University of Copenhagen

ESTHER GLAHN
University of Copenhagen

CARSTEN ROSENBERG HANSEN
Royal Danish School for Educational Studies

The purpose of the present article is to call attention to a number of shortcomings in much thinking about the "size of vocabularies". It also proposes that it would be sounder and more in keeping with reality to assume that vocabularies in reading are fluid and depend on the text read, on the reading strategies employed, and on the words the readers feel they know.

Since our assumptions are generated by a study of the size of the English vocabularies of Danish freshman students, we believe that we carry more conviction by presenting the original study in detail, with all its strengths and weaknesses, in order to show how it led us to more fundamental issues concerning the size of readers' vocabularies.

THE BACKGROUND OF THE VOCABULARY STUDY

The vocabulary study is one of several spin-offs from "Sprogtest" (i.e. "Language test"). "Sprogtest" checks Danish freshman undergraduates' global comprehension of written English by means of 45 multiple-choice questions in three English texts. The test

is used extensively, for Danish universities have no examinations of students' mastery of English, although many textbooks at university level are in English (Dollerup, Glahn & Rosenberg Hansen, 1980).

"Sprogtest" is used in freshman classes by individual teachers at Danish universities. Its purpose is to provide students with an anonymous, individualised, and objective assessment of their reading comprehension of English.

In our work with "Sprogtest", we found it pertinent to assess the size of the average Danish undergraduate's vocabularies. This simple enquiry led to the discussion and conclusion in the present article.

ASSUMPTIONS ABOUT VOCABULARIES.

There is a fair number of surveys and detailed studies of vocabularies (e.g. Anderson and Goodbody, 1981; Nation, 1983 ; Journal of Reading 29 (1986; special issue)). And the most important studies are regularly listed in e.g. Annual Summary of Investigations relating to Reading (IRA). In the present context, however, we need only summarize some points in the general thinking about vocabularies.

The size of vocabularies is frequently discussed. Thus Anderson and Goodbody (1981:96) cite researchers' suggestions that American sophomores have total vocabulary sizes of 15,000 words (Seashore in 1933); 19,000 (Kirkpatrick in 1907); of 60,000 (Seashore and Eckerson in 1940); of 85,300 (Gerlach in 1917); of 127,800 (Gilette in 1927); and of 200,000 different words (tokens) (Hartman in 1946), respectively (see also e.g. Nation (1983) chapter 10). Some of the discrepancies in the suggested sizes can be ascribed to different definitions of what constitutes "a word"; even so, the disagreement is striking.

Teachers, both of the mother tongue and of foreign languages, are understandably concerned with their students' vocabularies and with methods for increasing the number of words the students know (see e.g. Farr and Carey (1986)). There are numerous word lists, some of which are intended for the foreign learner of

English. This goes for e.g. Osgood and Richards' Basic with its 850 words that will do the work of over 20,000; Michael West's General Service List of 2,000 headwords; Bonger's list of 3,000 etc. (see Nation (1983) for further examples).

It is taken for granted that among native speakers of English "most people know all the very common words" (Anderson and Freebody, 1981: 101); and as very frequent words make up a large percentage of the running words in texts (see e.g. Anderson and Freebody, 1981; Nation, 1983), it is no surprise that frequency and frequency list are taken into account in language teaching:

For example, Thorndike-Lorge (1944) recommended what frequency bands the teachers' should concentrate on at different grades for teaching native speakers of English (Americans).

In English as a Foreign Language (EFL), the same idea explains why the 10,000 most frequent words from Thorndike-Lorge serve as a basis for teaching English in Sweden (Zettersten, 1979: 7). Noesgaard & Pedersen (1949) made a count of frequencies in four English primers for Danes; they correlated their lists with the first Thorndike frequency count (of 20,000 words), in order to enable Danish teachers to make a "scientifically sound" choice of English textbooks.

Today many teachers and researchers advocate that frequency lists should be supplemented with, e.g. specialised vocabularies:

Thus, the Americans Afflerbach, Allington and Walmsely (1980) suggest that U.S. citizens must learn the specialised vocabulary which people actually encounter in official application forms.

Many teachers of English as a foreign language share this view: Oppertshauser (1974) thinks that frequency bands must be modified for the needs of German students in view of the linguistic situations they will encounter; and Richards (1970), who believes French students should know 5,000 English words, also proposes to modify the automatic listings in frequency lists.

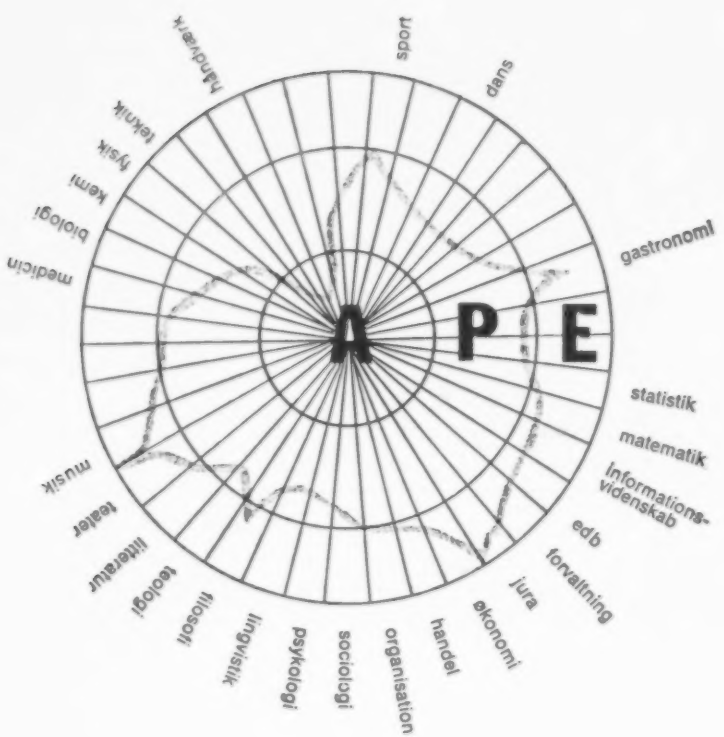
In brief: researchers and teachers agree that we cannot disregard frequency lists altogether, neither in teaching the mother tongue nor in teaching a foreign language.

We can therefore summarize the implicit and sometimes even explicit assumptions as follows: (a) frequency lists are not infallible yardsticks (e.g. Harris & Jacobson, 1975; Goodman &

Bird, 1985); (b) yet the very frequent words constitute a "core" (or "store") of words which all students, native speakers and foreign learners must/will learn and master. This basic "core" serves as a stepping stone for branching out into more specialised vocabularies concerning, e.g., our hobbies, interests, and backgrounds.

THE "CORE ASSUMPTION"

The above-mentioned "core assumption" is illustrated graphically by Hansen and Stetting (1977):



The figure is an abstraction: the outer ring shows the total linguistic capacity in the population; and the shading indicates the vocabulary of a lawyer ("jura") who is a specialist (with an E(xpert) language) on music and fine food ("gastronomi"). His generalised specialist language (P) reveals his interest in e.g. sports and philosophy and his total lack of interest in handicraft ("håndværk"); and, of course, he masters the general language (A), i.e. the syntax and the vocabulary of everyday communication, the "core" which we all know.

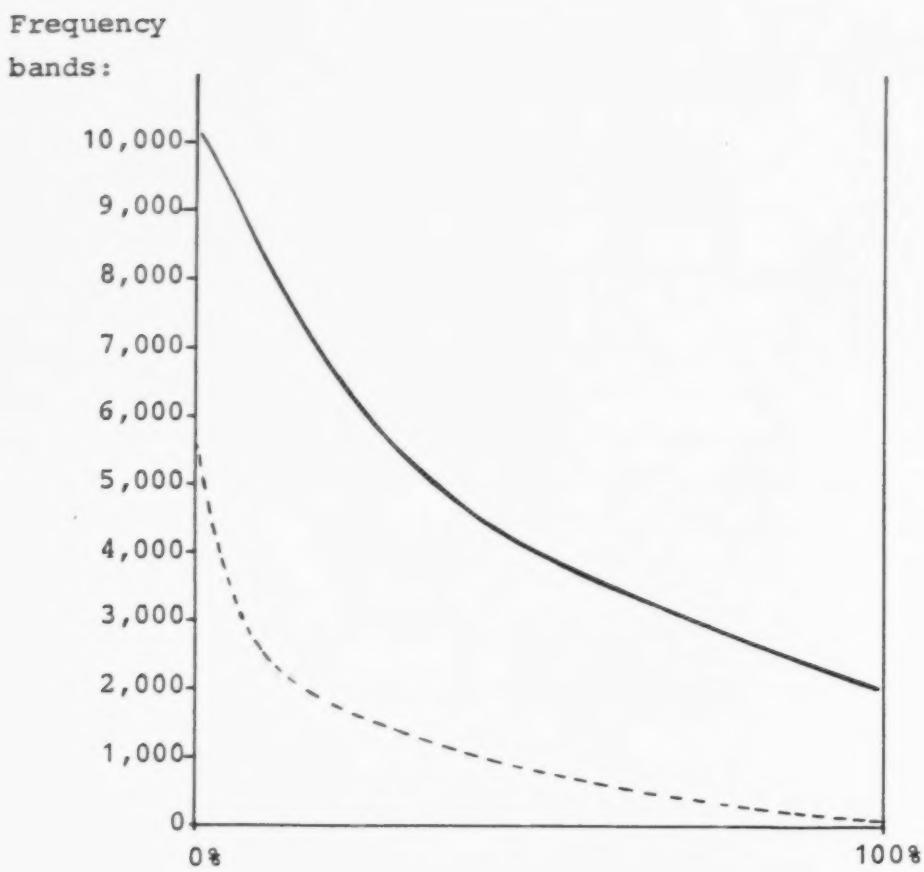
The "core assumption" is widespread among teachers - probably because their familiarity with a language is better than their students'; it is interesting that Brutten (1981) found that teachers were more inclined than students to pay attention to frequency when they identified the words they thought might be obstacles to comprehension.

Using frequency bands as our yardstick we can transfer the "core assumption" into two curves, one showing the vocabulary of a foreigner with a large English vocabulary, and another one with a fairly small vocabulary, respectively, as follows:

Curves showing the English vocabularies of two foreigners, against frequency bands.

——— a person with a "large" vocabulary.

------ a person with a "poor" vocabulary.



The graphs in the figure show how both foreigners know a large percentage or number of the "core" words. However, in the case of the more infrequent words, the learners' knowledge of English, their personal fields of interest, etc. will have an effect - but at all events the learners will know a diminishing number of words in the lower frequency bands.

The graph also obliquely illustrates another feature implicit in the core assumption: when we are dealing with large groups of learners, a high frequency word unknown to persons with large vocabularies should logically be even more unfamiliar to readers with small vocabularies. We must therefore assume that every time persons with large vocabularies do not know a word, even more people with small vocabularies will find the word unfamiliar.

VOCABULARY TESTING

Anderson & Freebody (1981) and Nation (1983) show how frequency bands affect the construction of vocabulary tests; some implicit "core assumption" affects vocabulary testing, as "frequency is a parameter which probably is very strongly related to probability that a word will be known" (Anderson & Freebody (1981: 101)).

Vocabularies are tested in numerous ways, for example in cloze tests (Johansson among Swedish students (1977)); pictures for comment (at the basic level in Denmark); multiple-choice questions, with either English or translated answers (the Cambridge Institute of Copenhagen; Arne Zettersten, 1979); etc. (For a detailed survey, see Farr and Carey (1986)).

Provided they are administered the same way to all learners, such tests fulfill useful diagnostic purposes. And yet they often have obvious weakness, in so far as the outcome is a combination of various factors, e.g. both receptive (passive) and productive (active) vocabularies; ability to understand syntax; pure guesswork, etc.

THE PARTICIPANTS AND THE PROCEDURE IN THE STUDY.

Thirty anonymous volunteer freshman undergraduates at the Department of English at the University of Copenhagen participated in the vocabulary study.

All participants answered four questions on their backgrounds; the answers are listed in appendix A.

Only the second question, viz., that on the readers' educational backgrounds calls for a comment here: attendance at the modern language line at the Danish "gymnasium" (i.e. A-level/high-school/Hochschule/lycée) implies that students had had from seven to nine years of English. At other lines in the 'gymnasium', the readers would have had from four to seven years of English. The "HF" is a line catering for mature students, at least in principle; students from this line will have done from five to eight years of English (a detailed description is given in Dollerup, 1981).

In addition, the participants in the vocabulary study were given six texts from the "Sprogtest" programme. The questionnaire, the instruction, and the six texts (whose order was rotated) were handed out in envelopes.

The instruction requested the undergraduates to underline all the words that they did not know, or did not understand from their immediate experience of the text.

We were aware that the instruction was ambiguous and might cover a wide spectrum: on the one hand some readers might underline any word they could not translate precisely in the context. On the other hand, other readers might underline only those words which they considered major stumblingblocks in their comprehension of the texts.

The texts were:

Switzerland: a newspaper article outlining a Swiss plan for the country's future development.

Miners: a newspaper article about British miners who worked to keep out water from the coal mines during a

strike (This article is given in Appendix B as an example).

Energy: a popular science article on the potentialities of geothermal energy, especially in the US.

Disasters: a popular science article on natural disasters (floods and droughts) which were ultimately caused by human exploitation of nature.

The Plague: the opening of Chapter xxxi in C.S. Lewis' Arrowsmith. It describes a plague spreading from China to the West Indies; and

The Special: the opening of Conan Coyle's short story The Lost Special where a man orders a special train for going to London.

The raw data are listed on the following page:

AN OVERVIEW OF THE DATA IN THE STUDY OF DANISH FRESHMAN STUDENTS' VOCABULARIES IN THE READING OF SIX ENGLISH TEXTS

TEXT	Total number of running words in each text ("type"), (divisible by five)	NUMBER OF DIFFERENT WORDS("tokens")			NUMBER OF WORDS UNDERLINED BY THE 30 READERS (Total, irrespective of their nature)	NUMBER OF DIFFERENT WORDS ("tokens") underlined
		Total	Number of proper names, abbreviations, and figures	Other words		
Switzerland	430	222	28 (mostly Swiss towns)	194	83	27
Miners	310	155	8 (names of mines)	147	98	27
Energy	710	365	45 (firms, localities and figures)	320	396	69
Disasters	420	251	30 (geographical)	221	216	37
The Plague	545	306	20 (geographical and fictional)	286	532	80
The special	570	269	21 (fictional and geographical)	248	260	46
Total in sample	2985(total number of words in sample)	1156 (no sum)	151 (India(n) occurs twice)	1050 (no sum)	1585 in sum	275 (no sum)

NOTE:"Totals in the sample" are not always identical with the sum of the figures from the six texts.

KODAK

ANALYSIS.

We knew that our undergraduates' vocabularies were not identical with those of native speakers, since large areas of everyday life in English are "blank spots" in foreigners' vocabularies. Futhermore, our students' "word knowledge" would derive from a variety of sources: teaching, personal reading, association with native speakers of English, etc. However, a cursory check of the thirty undergraduates' answers showed that the question of vocabularies was considerably more complex than we had facilely assumed. The data were studied from a series of different angles before we found the proper methods and explanatory models, but, of course, we leave out the cul-de-sacs in the below discussion.

PREPARATORY WORK.

The vocabularies studied are, then, the vocabularies of Danish readers reading an English text.

Virtually all separate typographical units in the text were defined as words (including figures and proper names).

The words were listed; homonyms; inflected forms (including plurals and different tenses); and derivations with the same obvious root (e.g. suffixes - but not prefixes) were placed together as one word, a token.

The words the readers underlined were incorporated in the table; all words which had been only partially underlined were listed in their complete forms. If readers underlined several words in succession, the words were listed individually.

RELIABILITY.

We have already touched upon the wide margin open to our anymous participants in their interpretation of "words they did not know".

This being so, our grading of our participants must necessar-

ily be relativistic. When we speak of our "best" readers in the subsequent discussion, it must be understood that this is only an operative definition, a convenient, stylistic short-hand for "readers who have underlined fairly few words in the six texts"; the five "best" readers underlined from 5 to 19 words in the whole sample (of 1156 different words).

Conversely, our five "poorest" readers are the five participants who underlined the highest numbers of words - namely, from 87 to 183 words.

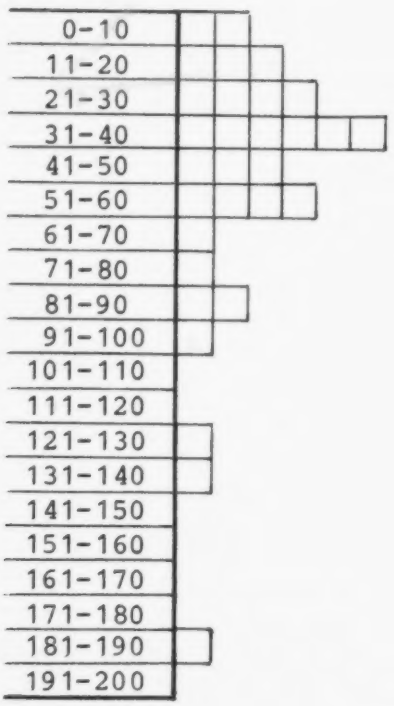
It can be objected that we should have used some kind of external control for assessing and validating our students's underlinings, e.g. a reading comprehension test. But, firstly, we doubt that our volunteers would have been willing to spend the time such a test would have required; and, secondly, no such test was available (apart from "Sprogtest" which would be selfdefeating as it includes three of our six texts in the vocabulary study).

The only means for external control is a comparison between the backgrounds of the five "best" readers and those of the five "poorest" readers (Appendix A). The duration of their stays in the English-speaking word, their educational backgrounds, and the number of books they had read seem to bear on the readers' familiarity with words: the two best readers - who underline 5 and 10 words, respectively - have both read more than 50 English books and spent more than a year in the English-speaking world. The importance of prolonged stays in the English-speaking world was also uncovered in Johanssons' study (1973) of Swedish undergraduates, so by and large there seems to be reason to assume that the greater the participants' familiarity with English, the lesser their inclination to underline words in the vocabulary study. There is, therefore, reason to believe that the underlinings are not completely relativistic although it this claim cannot be substantiated.

In another calculation, we listed the texts according to number of words underlined by all readers and compared the listings thus obtained (with corrections for variations in length in the texts) with the individual students' rankings. The readers were in agreement about which texts were "most difficult". We inter-

pret this as an indication that readers used some of the same criteria for underlining the words, and individually did so consistently throughout the texts. We believe the following figure ("Distribution of readers...") corroborates this hypothesis:

Distribution of readers according to the number of different words they underlined in the whole sample.



No. of words underlined:	No. of readers	Cumulated no. of readers
0-10	2	2
11-20	3	5
21-30	4	9
31-40	6	15
41-50	3	18
51-60	4	22
61-70	1	23
71-80	1	24
81-90	2	26
91-100	1	27
101-110		
111-120		
121-130	1	28
131-140	1	29
141-150		
151-160		
161-170		
171-180		
181-190	1	30
191-200		

□ = 1 reader

In this figure readers are grouped - or divided - according to the number of words they underlined. The point is that there is no regular petering off above the 60, or arguably, the 100 word-boundary. If our readers had applied completely identical criteria for underlining, we should have expected a regular levelling off - and this is not the case. We suggest this implies not only a knowledge of fewer words, but perhaps also the application of more complex criteria among the "poorest" readers, e.g. "I have not met with this word in this particular context", and "I would not know how to translate this word" (The strictest criterion is, of course, "I have not the vaguest idea what this unfamiliar word may mean").

DISCUSSION: THE READERS' VOCABULARIES.

The "Core assumption" and the Frequency Bands.

In this discussion we use the Thorndike-Lorge frequency-bands. Although there are major agreements between different counts in the highest frequency bands there are also variations in the order of the words in high frequency bands (e.g. Harris & Jacobsen, 1973; Dinnan, 1975).

Our choice of Thorndike-Lorge was determined by its comprehensiveness, i.e. it reaches far into the low-frequency bands, which therefore opens up the possibilities of including "rare" words in the discussion. Nevertheless, we think that the identity of the frequency count used is actually immaterial to our conclusions on questions of theory and principles.

It is generally accepted that the less frequent a word is, the smaller the chance that readers will know it. We can check this assumption with our data, ranking all words underlined according to the number of readers who underlined them.

A listing of one random word in each group looks as follows (with an indication of the frequency band in parenthesis):

1 reader: slope (2-3,000). 2 readers: emergency (3-4,000). 3 readers: hydraulic (10,000+). 4 readers: competitive (6-

7,000). 5 readers: sprawling (7-10,000). 6 readers: inconvenience (6-7,000). 7 readers: sinister (7-10,000). 8 readers: transpire (10,000+). 9 readers: crimson (3-4,000). 10 readers: pitching (2-3,000). 11 readers: spine (5-6,000). 12 readers: magma (not listed i.e. 30,000+). 13 readers: (a) stoop (2-3,000). 14 readers: sewer (7-10,000). 15 readers: immortelle (not listed). 16 readers: parched (7-10,000). 17 readers: pods (10,000+). 18 readers: bucaneers (10,000+). 19 readers: molten (6-7,000). 20 readers: semi-arid (arid: 7-10,000). 21 readers: scuppers (20,000+). 23 readers: corroborate (10,000+). 24 readers: sedateness (10,000+). 27 readers: incandescent (10,000+). 29 readers: seedie (not listed).

In general, the list corroborates the "core assumption": most readers are unfamiliar with frequent words, and fewer readers with infrequent words.

THE "GOOD" READERS.

We also posited that if the "core assumption" holds good, there must be more "poor" readers than "good" readers who will be unfamiliar with a specific word: every time a word has been underlined by a "good" reader, we must expect it to be underlined by even more "poor" readers.

Our five best readers had underlined 42 different words (tokens) 72 times. Only 2 of the 42 words failed to follow the pattern we showed above: in The Plague two "good" - but no "poor"-readers underlined the word "lather"; and four "good", but only three "poor" readers underlined "carreened". With these exceptions, the results also strengthen the core assumption.

The list of the words unfamiliar to our "best" readers deserves a closer scrutiny; in parenthesis we list the number of "good" readers that found any given word unfamiliar:

Switzerland: sedateness (1).

Miners: shotfirers (1); combustion (2).

Energy: fissures (1), harness (1), brine (1), crud (up) (1) feasible (1), sulfur (1); ample (2), (non-)corrosive (2), molten (2); fiscal (3), rudimentary, (3).

Disasters: devastating (1), prodigious (1), squander (1); parched (2), semiarid (2), rampaging (2); inundated (3).

The Plague: clattering (1), pods (1), hibiscus (1), buccaneer (1), berth (1); vigilant (2), boisterously (2), sewer (2), lather (2), incandescent (2), scuppers (2), immortelle (2); carreen (4), seedie (boy) (4).

The Special: stoker (1), ascertain (1), stoop (1), spine (1), dispatch box (1); oscillation (2); corroborate (4).

Most of these words are rare: "ample" - the most common and frequent word underlined by our "best" readers is in the 3-4,000 word band; "combustion" and "molten" in the 6-7,000; and "parched" in the 7-10,000 word band. The majority of the unknown words are in the 10-20,000 word band, with "scuppers" and "rampaging" in the 20,000+ band. And, as mentioned, "immortelle" and "seedie (boy)" are not listed by Thorndike-Lorge at all.

Conversely, if the words in the texts serve as the point of departure the five "best" readers had no problems with numerous words in the 10,000+ range, e.g. collated, deforestation, ecological, geological, hectar, jumbo jet (set), nuclear, reforest, technological, savanna(h), round-the clock, supplemental, unsparing, thermonuclear, overblown, supercargo.

Some of these words are undoubtedly more common today than when the corpus of the Thorndike-Lorge count was written, e.g. nuclear. Even so, the best readers know many highly infrequent words: their vocabularies are very large, and not confined to words from their own specialist areas. It is true that some of these words, e.g. hectar, savanna(h) also exist in Danish. But if we uncritically accepted that Danish readers would know English words which looked like Danish ones we would miss a point: these

words are not very frequent in Danish either, so the impression that some readers have large receptive vocabularies is not weakened.

THE FIVE "POOREST" READERS.

We would expect our poorest readers to know only "core-words" and then only odd words above a certain boundary (which would, in turn, depend on the readers' knowledge of English). As mentioned, our "poorest" reader underlined 187 words. Among the unfamiliar words were e.g. current (1-2,000 word band); acknowledge in the 3-4,000 word band; complex (5-6,000), etc. But, curiously, words like available, code, and economy (3-4,000 word band); dilemma (10,000+); depopulate (20,000+), and many similarly infrequent words were not underlined.

ALL THIRTY READERS.

Affair, bright, forest, c(ent), and guard in the 0-1,000 word band were each underlined by only one reader. So were bore, bound, current, firm, flat in the 1-2,000 word band. In the 2-3,000 band attach, commit, and depth were likewise unknown to one reader each - only application was unfamiliar to 5 readers. In the 3-4,000 word band apparent was unknown to one reader: two readers underlined available, contribution, decrease, emergency: and no less than 12 readers indicated that ample was unknown to them. However, if we look at the texts in another way, the list of words from the low frequency bands unfamiliar or unknown to only one of the thirty participants looks as follows:

3-4,000 word band: amaze, chapel.
4-5,000 : barrier, cargo.
5-6,000 : banana, breathless, complex.
6-7,000 : balcony, bamboo.
7-10,000 : breakdown, annal, conservation, comical, dependence, first-class, fragile, market-

place, phenomena, rainfall, sensational, ski, skipper, smear, spokesman, spontaneous, underlying, urban.

10,000+ : bazaar, centre, dilemma, efficiently, ensure, exotic, fantastically, geysir, ineffiecient, inexplicable, middle-aged, monsoon, deforest, depopulate, geological, nuclear, overblown, reforest, supplemental, technological, periodically, physique, potentially, seasonal, second-class, turbine, upstream, washerwoman.

30,000+ : (i.e. not listed): hectar, round-the clock, breakthrough.

DISCUSSION.

The "core assumption" appears to hold good as virtually no Danish readers of English at an advanced level meet with unfamiliar words in reading below the 3-5,000 word boundary.

The exact boundary can, however, not be defined. Even if we had established it, we could not claim that it would aply to all learners of EFL: in other words, we cannot and will not argue that all learners of EFL must know any specific number of words in order to manage.

In addition, there is an equally important result: many undergraduates appear to know even infrequent words, and this cannot be explained by simply combining the "core assumption" with frequency bands: many of the words discussed would be very infrequent in any general frequency count of the English language - hence it did not matter that we used Thorndike-Lorge.

VOCABULARIES AND READING STRATEGIES.

As mentioned, the "Sprogtest"-programme comprises other studies than the vocabulary study, including an introspection study where 28 other readers - 7 undergraduates and 21 students at the modern language line at the gymnasium - reported on their

reading and test-solving techniques during the reading (Dollerup, Glahn & Rosenberg Hansen, 1982).

This particular study leads us to suggest that the "core assumption" should be supplemented with reading and decoding strategies. This would explain why our readers had fewer difficulties with low-frequency words than expected.

However, we cannot assess the influence of reading and decoding strategies in the vocabulary study; this would require an introspection study and a vocabulary study at the same time and we simply do not have the resources for such a study.[1]

Therefore we cannot document, e.g., "common sense", and background knowledge; and we cannot hypothesise the influence of some strategies, e.g. "reference to syntax and grammar" in the vocabulary study with its primitive underlinings.

Even so, there are several strategies for decoding English words in the introspections study which suffice for us to make our point with the present material: E.g.

1. Etymological, morphological, and/or (transparent) semantic decoding, i.e. readers use:

1a. Components of words they know from another language (mostly Latin): Text: "Decreasing the inconvenience' Reader's comment: I don't know how to translate 'decreasing'. Then I think of Latin 'convenio'..."

1b. Components from English words familiar to the readers.

1c. A knowledge of a Danish word which looks more or less like the one read: for example, the English word flood (inundation) was often taken to mean 'river' which translates as Danish "flod" (a so-called 'false friend').

2. Translation into Danish. The speech cited at 1a. illustrates this strategy which applies to both passages and words (compounds).

3. Context: e.g. "I have seen these words before, but I do not know what they mean: when it says 'the first carriage was solely...' this must means 'only'. I go for the first answer to the multiple-choice question because I skim the text. It

says that the carriage has only first and second class compartments."

We suggest that these (and other) strategies explain why the students' know low-frequency words in the vocabulary study:

Semantic/morphological/etymological decoding from foreign languages (1a) would explain why readers have little problem with e.g. 'depopulation' - and with hybrid forms such as 'deforest' and 'reforest'. This assumption is also borne out by the fact that some readers underlined only part of a compound, e.g. arid in semiarid, and corrosive in noncorrosive.

A knowledge of some of the English components of the words (1b) must overlap with translation in our "vocabulary study", but the undergraduates' knowledge of the following compounds and derivatives are - it seems - easiest to explain in this fashion: breathless, breakdown, dependence, rainfall, underlying, inefficient, inexplicable, overblow, seasonal, upstream, breakthrough, round-the-clock.

Similarily, Danish words (or Danish loan words)(1c) may influence the readers' knowledge of the following words: chapel, barrier, banana, complex, balcony, bamboo, annal, comical, phenomena, sensational, ski, skipper, bazaar, center, hectar, etc.

Translation into Danish of the individual components (2a) would explain the readers' familiarity with e.g. firstclass, market-place, middle-aged, washerwoman.

Context (3) was used extensively by readers in the introspection study - as it is by native speakers of English (Nagy, Herman & Anderson 1985). Nevertheless,it is not well documented in our vocabulary study. However, it is suggestive that the word "ample" in the 3-4,000 word band stands out as word unfamiliar to twelve of the thirty readers, including two of the best. The word appears as a subtitle in Energy and its meaning is not easily decoded from the context. In Miners, context also seems to affect the underlinings: the first time the word pit appears in the text, seven readers underlined it - the second time, only one reader did so. There are presumably two reasons: some readers decoded the words meaning from the context, and

others thought that since they had already once indicated the word was unfamiliar, it was unnecessary to repeat this information. It is also tempting to assume that the readers' familiarity with e.g. fragile and smear depends on context; these words appear in the sentences "Because the semiarid region is ecologically fragile" and "the slaves came on the tyrant lord and straightaway the lather was smeared with blood", respectively. In both cases, however, influence from other languages, viz., French (fragile), and Danish (indsmøre) may count as well.

In brief, we suggest that if reading and decoding strategies are taken into account, they will explain the readers' understanding of infrequent words.

One last point - also mentioned by Anderson & Freebody (1981) and Nation (1983) must be made, viz., that the concept of knowing a word is problematic. From our sample, it seems as if one strategy is to get a hazy idea of what a word means, assess that it is fairly unimportant, and then accept this vague impression as "familiarity"; thus only half the readers underline the word immortelle, presumably because it occurs in the sentence: "the immortelle that fills the valleys with crimson". The sentence signals that immortelle is a kind of red large flower, and in the large context, it serves only to give flavour to the description of a tropical island: readers have an ability to discriminate the importance of words (Brutten, 1981).

CONCLUDING REMARKS.

The readers who participated in our vocabulary study are well above the average learners of EFL; this is so partly because foreign language teaching figures prominently in the Danish school; partly because of the pervading "English" influence by means of an influx of foreign language material (films, music, etc.), a high foreign trade, and tourism; and partly because Danish and English are closely related languages. Nevertheless we suggest that as readers they shed light on vocabularies in the reading process, not only of texts in a foreign language but in the mother tongue as well.

Above all, the study shows that we must take care to define what type of vocabulary we are talking about. In some contexts, it may suffice to use terms such as receptive (passive), and productive (active); but frequently this is not enough: we must specify the special function of the vocabulary, or the situation in which it is used or in which it has been charted.

Frequency lists will continue to be used. However, in addition to the shortcomings and limitations in terms of extensiveness of the corpora and their place in time and space, frequency lists are static, whereas the language and the words we meet with in "real life" are dynamic. Frequency counts may, or may not, by accident or by intention, list nonce-words and words which exist for a short time only, although such words may be understood by many people, for instance, because these words follow ordinary rules for creative word formation, or are understandable from the social and linguistic context in which they are used. Frequency lists have these limitations in common with even the largest dictionaries; it is impossible to keep track of the emergence and disappearance of all linguistic entities. But in a discussion of somebody's vocabulary we cannot disregard such words - they are part of the process of communication.

We suggest that in reading, we are not dealing with a static entity when we speak about a vocabulary but a changing and fluid mass.

There is a core of words, a word knowledge, which centres around the most frequent words in the language and the size of which may vary with readers' personalities and backgrounds. This word knowledge is, we suggest, relatively - but not completely - stable, and its size can be estimated, with the limitations imposed by the methods used and the definitions of vocabularies employed. <u>But this word knowledge is only part of a readers' receptive vocabulary</u>.

Another part of the vocabulary consists of <u>the strategies that individual readers use for decoding words and for gaining an overall comprehension</u>. This has been touched upon by others. Thus, Arnauld (1984) cites Denninghaus as having used the term "potential vocabulary" about words hypothetically known to

learners. Nagy & Anderson (1984) suggest that knowledge of infrequent words increases with exposure to language, and refines this in Nagy, Herman & Anderson (1985) to an ability to learn by context. We wish to stress, however, that (a) the strategies are not identical with a learning process but that the words are understood and known in one particular context and perhaps only momentarily, and (b) that this applies to reading. We do not preclude that this approach applies to other situations as well, but leave this problem for others to solve.

A third component of a vocabulary is the text which is actually being read: it is only in the reading of a text that the strategies and the word knowledge can interplay. To be explicit: there are words which an individual reader will meet with and immediately understand only once in a lifetime.

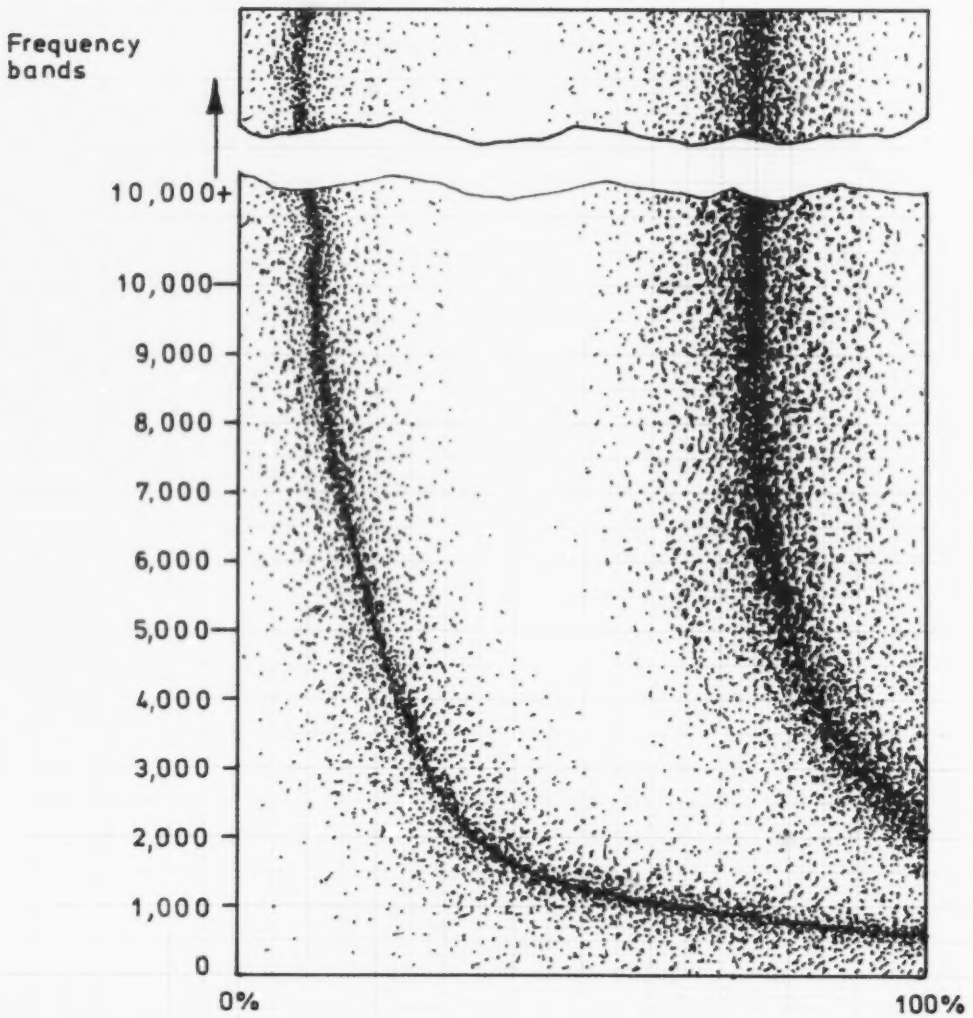
In summary: readers' vocabularies in the reading process consists of (a) a "word knowledge /store", of (b) strategies for decoding words, and of (c) the special linguistic context. It implies that individual vocabularies in reading exist instantaneously, and that they are, in effect, fluid entitities which change every time they are generated by the reading of specific texts. Vocabularies differ not only in time but also from text to text with the same reader.

This leads to a different figure for individual vocabularies than the one we previously showed, namely the below sketch of individual receptive vocabularies in reading viz.

A sketch of individual receptive vocabularies in reading.

reader with a large core vocabulary and many strategies.

reader with a small core vocabulary and few strategies.



In this figure the left hand column indicates the frequency bands. It includes all words in a specific language, even those not listed in the most comprehensive dictionaries: therefore we leave the upper limit open (which does not mean that vocabularies are infinite).

In adding the readers' reading and decoding strategies, we suggest that poor readers with few strategies at their disposal will know fewer words in any given frequency band than the good readers'; yet they will still know some very rare words.

These conclusions tally with what we find in "the real world": it is true that beginners' books may be edited for beginners' usage. But most books - textbooks and otherwise - are not edited at all. They make no allowance for those who find it hard to read their mother tongue, and show no special toleration of the foreigner who must read them. There is no reason to doubt that the total number of words in printed school English in the US is somewhere around 88,500 distinct words - or more - and that the average American student will encounter at least 5,000 distinct word families (Nagy & Anderson, 1984).

No real-life text is guided by core-words. On the contrary, it is guided by its contents: and for adequate coverage of the contents, the author may have to use many low frequency words (Goodman and Bird, 1985).

And, finally, as we have pointed out, the total linguistic potential in a language community allows for the inclusion of completely new words, neologisms, ad-hoc-formations etc. which are understood by others (in other words: language is dynamic - like the individual vocabulary).

The results indicate that the importance of vocabularies should not be overlooked; there should be a conscious instruction in the rules of word formation and of word derivation.

And most of all: reading strategies should be taught as an integral part of these activities.

In vocabulary testing it should more readily be acknowledged that frequency lists may tell some part of the truth and a useful one at that - but sometimes they are a far cry from the whole truth.[2]

NOTES

1. We did in fact undertake a supplementary investigation along the same lines as the "vocabulary study" with the edited texts used in The "Sprogtest" proper. 28 readers participated. However the result was the same as in the vocabulary study which we had chosen for the above incisive discussion since it deals with the original, authentic texts.

2. We wish to thank all the readers who participated in the present study. We are indebted to Ethel Ussing for her unfailing help in setting up the material for "Sprogtest", and for the present vocabulary study; and to Anette Andersen and Marlene Barner for having typed this article. We are grateful to our assistants over the years, viz., Benedicte Holbak, Birte Kristensen, Karin Sigurdskjold and Eva Schaumann.

REFERENCE.

Afflerbach, Peter P., Richard L. Allington, and Sean A. Walmsley. (1980) A Basic Vocabulary of U.S Federal Social Program Applications and Forms. Journal of Reading 23, 332-336.

Anderson, Richard C. and Peter Freebody. (1981) Vocabulary Knowledge. In Guthrie, John T. Ed. Comprehension and Teaching: Research Views (Newark: IRA, 77-117).

Arnauld, Pierre J.L. (1984) A Practical Comparison of five types of vocabulary tests and an investigation into the nature of L2 lexical competence (paper read at AILA, Bruxelles, 5-10 August, 1984).

Brutten, Sheila R. (1981) An Analysis of Student and Teacher Indications of Vocabulary Difficulty. RELC-journal, 12, 66-71.

Dinnan, James A. (1975) A comparison of Thorndike-Lorge and Carrol prime frequency word lists. Reading Improvement, 12 , 44-46.

Dollerup, Cay, Esther Glahn, Carsten Rosenberg Hansen. (1980) Some Errors in Reading Comprehension. In Faber, H. von Ed. Leseverstehen im Fremdsprachenunterricht (Munich: Goethe Institut).

Dollerup, Cay. (1981) Studies in the Major Modern Languages (English, German, French) at University Level in Denmark by 1980/81 (ERIC ED 203 681).

Dollerup, Cay, Esther Glahn, Carsten Rosenberg Hansen. (1982) Reading Strategies and Test-Solving Techniques in an EFL-Reading Comprehension Test: a Preliminary Report. Journal of Applied Language Study, 1, 93-99.

Farr, Roger and Robert F Carey. (1986) Reading: What can be measured (2nd ed. Newark: IRA).

Goodman, Kenneth S. and Lois Bridges Bird. (1984) On the Wording of Texts: a Study of Intra-Text Word Frequency. Research in the Teaching of English, 18, 119-145.

Hansen, Inge Gorm, and Karen Stetting. (1977) Specialsprog. In Glahn, Esther and Leif Kvistgaard. Eds. Fremmedsprogspædagogik (Copenhagen: Akademisk forlag).

Harris, Albert J, and Milton D. Jacobson. (1973-74) Some Comparisons between Basic Elementary Reading Vocabularies and Other Word Lists. Reading Research Quarterly, 9, 87-109.

Johansson, Stig. Reading comprehension in the native and the foreign language: on an English-Swedish comprehension index. (1977) In Zettersten, Arne. Ed. Papers on English Language Testing in Scandinavia (Copenhagen: Anglica et Americana 1, 43-58.

Journal of Reading (IRA). (1986: no. 7. Special Issue on Vocabulary).

Nagy, William E. and Richard C. Anderson. (1984) How many words are there in printed school English? Reading Research Quarterly, 19, 304-330.

Nagy, William E. and Patricia A. Herman, and Richard C. Anderson. (1985) Learning words from context. Reading Research Quarterly, 20, 233-253.

Nation, I.S. Paul. (1983) Teaching and Learning Vocabulary (Victoria University of Wellington: English Language Institute).

Noesgaard, A. og Vagn Pedersen. (1949) Hyppighedsundersøgelser over Engelsk som Fremmedsprog (fire begynderbøger). (Copenhagen: Fr. Bagge).

Oppertshauser, Otto. Absolute oder relative Haufigkeit? Wort-Statistik als Hilfsmittel zur Aufstellung eines verbindlichen Mindestwortschatzes fur den Englischunterricht im Sekundarbereich I. (1974) Praxis des neusprachlichen Unterrichts, 31, 42-52.

Richards, Jack C. (1970) A Psycholinguistic Measure of Vocabulary Selection. IRAL, 8, 87-102.

Thorndike, Edward L. and Irving Lorge. (1944) The Teacher's Word Book of 30,000 Words. (New York: Columbia University).

Zettersten, Arne. (1979) Experiments in English Vocabulary Testing (Malmö: Liber Hermods).

APPENDIX A

BACKGROUND OF PARTICIPANTS IN THE VOCABULARY STUDY

VARIABLE		Number among all thirty participants	Number among the five best participants	Number among the five poorest readers
AGE	Years			
	18	3	0	1
	19	6	1	0
	20	9	2	3
	21	2	0	0
	22	1	0	0
	23	2	0	0
	over 23	7	2	1
	Total	30	5	5
ACADEMIC BACKGROUND (simplified)	Modern languages at the gymnasium	14	3	1
	Other line at the gymnasium	9	0	2
	HF	7	2	2
	Total	30	5	5
DURATION OF STAYS IN ENGLISH SPEAKING WORLD	No stays	8	1	2
	less than a month	8	0	3
	1 month - 3 months	1	0	0
	3 months - 1 year	7	1	0
	More than a year	5	3	0
	English home	1	0	0
	Total	30	5	5
NUMBER OF BOOKS IN ENGLISH READ	1 - 2 books	2	0	2
	3 - 9 books	7	1	2
	9 - 50 books	15	2	1
	more than 50 books	6	2	0
	Total	30	5	5

APPENDIX B

VOLUNTEER MINERS HAVE KEPT COAL-FACES SAFE

Voluntary safety work carried out by miners during their strike means that all coal faces are expected to resume production when the strike ends.

After the last strike 25 out of about 800 coal-faces were lost but co-operation between miners and management has ensured that the pits have been kept safe during the present strike, now in its 23rd day.

Apart from deputies and overseers who have reported for duty each day, many miners have volunteered to help in taking safety measures with their union's approval.

A Coal Board spokesman said yesterday: "The vast majority of coal faces are standing up well. Precautions were taken in the light of experience gained during the last strike.

"Each pit has its own peculiarities and weaknesses and it is the men who know them, at local level, who have gone down to see what has been happening, and where necessary to take action."

Precautions taken

The main trouble has been that in some pits the roof supports, which are hydraulically controlled, have locked solid. There have also been some breakdowns in pumping machinery used to keep certain sections clear of water, but men have gone down to repair and replace the machinery where necessary.

At the start of the strike, precautions were taken at certain pits against spontaneous combustion by sealing off sections likely to be affected. In many cases the management has been working seven days a week and the union has kept a committee available round-the-clock to instruct members whose help has been needed in emergencies.

At Blaenant, South Wales, teams of miners worked three shifts day and night for a week to reduce water levels after a major pump breakdown. At Bedlay, Glasgow, men on strike worked with members of the National Association of Colliery Overseers, Deputies and Shotfirers to clear a big fall at one of the coal-faces.